Copy number variations in multiple sclerosis

Anna-Maija Sulonen

Institute of Molecular Medicine Finland, FIMM

The National Institute for Health and Welfare

5.3.2012

Research thesis

Supervisor:

Janna Saarela, MD, PhD, Docent

Institute for molecular medicine Finland, FIMM

UNIVERSITY OF HELSINKI

Faculty of Medicine

anna-maija.sulonen@helsinki.fi

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET

Tiedekunta/Osasto – Fakultet/Sektion – Faculty	Laitos – Institution – Department
Faculty of Medicine	

Tekijä – Författare – Author

Anna-Maija Sulonen

Työn nimi – Arbetets titel – Title

Copy number variations in multiple sclerosis

Oppiaine – Läroämne – Subject

Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä -Sidoantal - Number of pages
	March 2012	18+4

Tiivistelmä – Referat – Abstract

Current SNP genotyping arrays for genome wide association studies (GWA) are used to study both SNPs and CNVs. We conducted a GWA study for multiple sclerosis (MS) using 68 distantly related MS cases from Finnish subisolate with high risk for MS in Southern Ostrobotnia and the Illumina HumanHap300 chip. We identified 106 CNV regions in the MS cases with QuantiSNP v1.0 software, and confirmed the data by visual inspection of the BeadStudio intensity data. Common pathways were searched for genes mapping to or near the identified CNV regions. We found one pathway involving five genes, *ERBB4*, *NRG3*, *DLG2*, *UTRN* and *LARGE*, regulating oligodendrocyte development and survival. An in house-built genotyping method was used to genotype three of the deletions in the pathway in a large nation-wide sample set of both cases and controls. Evidence for association was not found.

(139 words)

Avainsanat – Nyckelord – Keywords

DNA Copy Number Variations; Multiple Sclerosis; Genome-wide Association Study; Genetics

Säilytyspaikka – Förvaringställe – Where deposited

Muita tietoja – Övriga uppgifter – Additional information

1 Introduction

Multiple sclerosis (MS) is a chronic inflammatory disease of the central nervous system, characterized by demyelination and axon damage (1). Common symptoms of MS include paresis and paresthesia, disturbances of gait and coordination, loss of vision and extreme fatigue. In most cases, there is complete recovery from the first symptoms, but later the disease progress in a relapsing-remitting manner. Typical age of onset is 20-40 years, making MS is the most common severe neurological disease of young adults. MS is a complex disease with both environmental and genetic risk factors. In Finland, prevalence of MS is approximately 1:1000, and twice as high, 2:1000, in the Southern Ostrobothnia (2). This difference is thought to be explained at least partly by genetic risk factors enriched in the subisolate. To date, no Southern Ostrobothnian specific genetic mutations causative for the disease have been identified.

Recent studies in MS genetics emphasize the role of the immune system and enlighten the pathology of MS as an autoimmune disease. Variants in genes related to various immune cell functions, such as the HLA locus, IL2RA, IL7R, CLEC16A, CD58, TNFRSF1A, IRF8, STAT3 and TYK2, have been associated with MS disease (3-7), and these findings have just been further verified by a large multicenter collaboration in a genome wide association study (GWA) (8). Yet, these variants explain only a fraction of the inheritance of the MS disease (8). These studies have been carried out by studying single nucleotide polymorphisms (SNPs), which are common in healthy individuals as well. In 2004, Iafrate *et al* published a novel form of genetic variation, further defined by Redon *et al* and McCarroll *et al* in 2006: variation in copy number of relatively short DNA fragments (copy number variations, CNVs) (9-11). CNVs are believed to explain as much of the variance between individuals as SNPs (10-11). Since then, the role of CNVs in disease susceptibility has been widely studied.

Current SNP arrays for genome wide association studies are used to study both SNPs and CNVs. We conducted a GWA study for multiple sclerosis using 68 distantly related MS cases from Finnish MS high risk sub-isolate of Southern Ostrobotnia and the HumanHap300 chip from Illumina. Although the used chip has not been designed for monitoring CNVs, we wanted to take a first look of CNVs in MS. We assessed the frequency of the identified CNVs by comparison to public databases, and evaluated their possible role in the etiology of MS with pathway analyses. Furthermore, we applied an in-house developed CNV genotyping method to carry out genotyping of three deletions in high-through put manner in a nationwide sample set of more than 700 cases and 1000 controls.

Results of this study have been previously reported in Jakkula, *et al*, 2010 (7).

2 Material and methods

2.1 Samples

MS cases, which had either both parents born in the MS high risk region in the Southern Ostrobothnia or one isolate-born parent and family history of MS, were selected for the study. The diagnosis of MS was determined using Poser's criteria (12). Peripheral blood samples were collected during 1998-2002 and DNA was extracted using conventional methods at National Institute of Health and Welfare. The birthplaces of cases and their parents were collected and confirmed from national registries. After reviewing the birth places and names of the parents, population records were used to gather genealogical data up to 5-18 generations back. Two large pedigrees were identified, which date back to 15th century (Figure 1). 68 MS patients were ascertained to origin from the isolate and were selected for genome wide SNP genotyping.

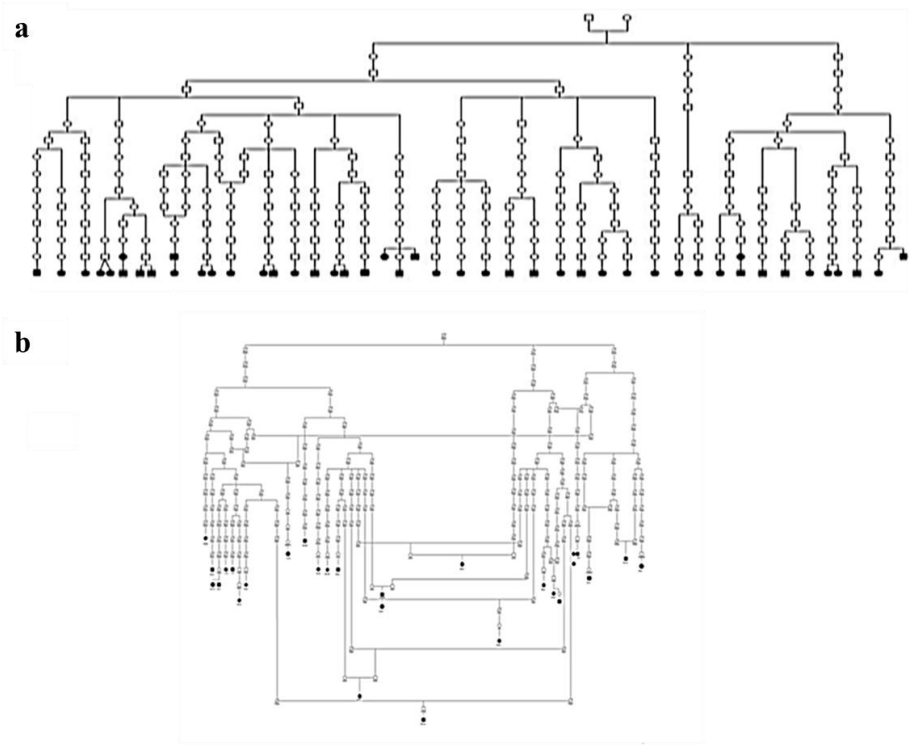

Figure 1. MS pedigrees from the high risk region of Southern Ostrobothnia. Although profound interrelatedness does exist between the family members, only the shortest connections are presented. All family members are not depicted for illustration purposes. Figure modified from Jakkula, *et al*, 2010.

2.2 Ethical aspects

Sample collection and the study design have been approved by the Helsinki University Hospital Ethical Committee of Ophthalmology, Otorhinolaryngology, Neurology and Neurosurgery (permit 192/E9/02). Informed, written consent was obtained from all patients participating to the study. Statistical analyses were performed as a large combined dataset and interpreted only as such. No individual genotypes or other personal results were reported back to the patients or their family.

2.3 SNP genotyping

Illumina HumanHap300 SNP genotyping chip consist of arrayed 50 base pair long DNA probes complimentary to the known reference sequence next to the 317,503 SNPs to be genotyped. Sample DNA is hybridized on the array, followed by single nucleotide extension with labeled A, C, G and T nucleotides and imaging of the chip. Illumination of the different nucleotides (specified A and B allele signal intensities for either of the alleles for each SNP) are measured, and genotypes are assigned with automated algorithms. The 68 MS cases chosen for the study were genotyped at the Broad Institute, Cambridge, USA, with Illumina HumanHap300 chip according to manufacturer's protocol (Illumina, Inc., San Diego, California, USA). SNP ID's and positions were mapped to NCBI human genome build 35 (hg17).

2.4 CNV estimation

Identification of copy number variations from SNP chip data is based on the signal intensity of the SNP genotyping probes ("Log R Ratio") and the allelic signal intensity ("B Allele Freq") of the SNPs on the length of the chromosomes. For the loci where a chromosomal region is duplicated, signal intensities of the SNP probes are multiplied when compared to that of the chromosomal regions with normal, haploid copy number, in proportion to the number of times the region is duplicated. At the same time, the allelic signal intensities of the heterozygote SNPs in the duplicated regions shows typical clustering: for triplicated regions, genotype clusters for AAB and ABB genotypes can be

identified, for quadrupled regions, genotype clusters for AAAB, AABB and ABBB can be identified, and so on. Figure 2(a) shows an example of a triplicated locus. On the other hand, for the loci where a chromosomal region is lost and thus the locus is present only in one chromosome, the signal intensities of the SNP probes are lower from that of the signal intensities of the SNPs in the loci with normal copy number. At the same time heterozygous SNPs are lost from the region (Figure 2(b)). If both copies of the locus are lost from the chromosomes, resulting in a copy number of 0, signal intensities of the SNP probes fall to immeasurable and the allelic signal intensities show typical scattering across all possible values.

For this study, CNVs were estimated for the 68 cases using QuantiSNP v1.0 software (13). The local G and C base content of the sequence next to the genotyped SNPs affects the probe hybridization and the intensities subsequently, so the GC correction option of QuantiSNP was used. As recommended by the software authors, only CNVs with a log Bayesian factor >10, indicating reliable CNV calls, were included in further analyses. All CNV calls were assured visually from the Illumina BeadStudio software's allelic intensity ratio ("B Allele Freq") and normalized total intensity ("Log R ratio") scatter plots (Figure 2). All CNVs with discrepancies between the automatic calls and visual estimation were excluded from further analyses. Additionally, centromeric regions were excluded, and only the CNVs with three or more SNPs were taken into account. PennCNV software (2007Nov13 version (14)) was also used, but as this seemed to yield more false positives when the results were visually estimated, its results were not used in further analyses. Copy number variations were regarded as same events between the samples if they overlapped in chromosomal position. Copy numbers 3 and 4 were regarded jointly as "gain" events and copy numbers 0 and 1 as "loss" events.

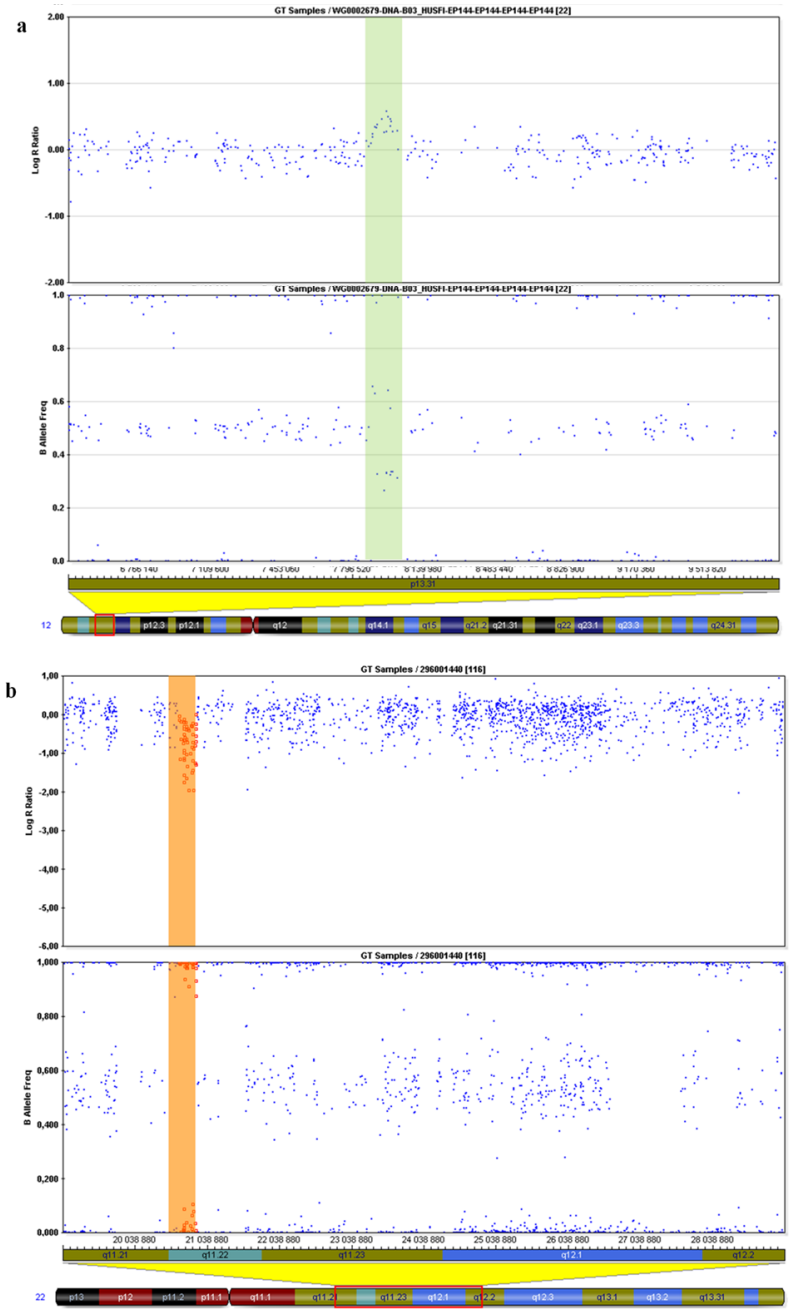

Figure 2. Examples of visual estimation of QuantiSNP CNV calls. Screen captures from Illumina BeadStudio showing a typical duplication (a) in the genomic region highlighted with green in chromosome 12 with elevated SNP probe signal intensity (Log R Ratio, upper panel) and clustered allelic signal intensities (B Allele Freq, lower panel) of heterozygous SNPs, and a typical deletion (b) in the genomic region highlighted with orange in chromosome 22 with a drop in SNP probe signal intensity and no heterozygous SNPs. Blue dots represent individual SNPs.

2.5 Pathway analyses

Pathway analyses were generated through the use of Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com). Genes mapping near or into the identified CNVs were given as an input feed. Pathway growing –option, allowing addition of 5 molecules, and connect-option, connecting the input and added molecules in the analyses, were used. All output pathways were critically studied from literature.

2.6 CNV genotyping

The copy number variations mapping to a joint pathway (see Results) were further genotyped in a larger set of 703 MS cases and 1051 controls from all over Finland. For genotyping of the deletions in *DLG2*, *ERBB4* and *NRG3* promoter, we used the in-house built method (InGoLF, Instant Genotyping of Lost DNA Fragments) described in Sulonen, *et al*, 2009 (15). In SNP based copy number variation estimation the true break points of the CNV lie somewhere between the last undeleted or unduplicated SNP and the first deleted or duplicated SNP. We sought for the break points by designing primers (Primer3Plus (16)) to these "grey zones" sequentially, approximately 1000 base pairs from one another, forward-primers to the 5'-end of the estimated deletion and reverse-primers to the 3'-end. As we then performed PCR with different primer pair combinations and re-sequenced the shortest product from the individuals with the estimated deletion, we found the exact break points of the deletion.

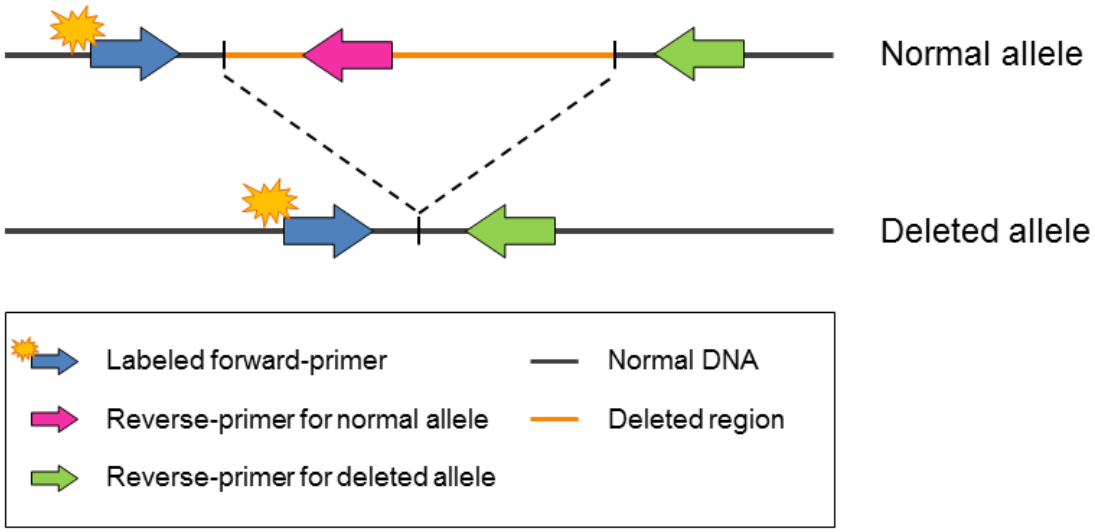

Figure 3. Principle of the InGoLF-genotyping. Schematic presentation of the primer design for different alleles for the genotyped deletions. PCR product over the deleted region is too long to be efficiently amplified in the genotyping PCR. PCR products from the different alleles are of different sizes, and thus the genotypes can be separated by gel electrophoresis. Figure modified from Sulonen, *et al*, 2009.

For the InGoLF-genotyping we designed six different PCR-products across these break points, using nine primers: one product from the normal allele and one from the deleted allele for each of the three deletions. Schematic presentation of the assay design is given in Figure 3. All deletions had one labeled primer (FAM, VIC and NED), used jointly with the two different primers for the different alleles. PCR-products from the two alleles for each deletion were of different sizes, ranging from 155 base pairs to 565 base pairs. Table 1 describes the primer design and lengths of the PCR products. PCRs for the deletions in *DLG2* and *NRG3* were done in one multiplex reaction and PCRs for the deletion in *ERBB4* in its own multiplex, each in the following conditions: 10 ng of DNA, 200 nM of the labeled forward-primer and 300 nM of the allele specific reverse-primers for the DLG2-deletion and ERBB4-deletion, 140 nM of labeled forward-primer and 200 nM of allele specific reverse-primers for the NRG3-deletion, 0.55 IU of AmpliTaq Gold PCR polymerase (Applied Biosystems, Foster City, CA, USA), 1x AmpliTaq reaction buffer (Applied Biosystems), 200 µM of dNTP, 1.5 mM on $MgCl_2$ and ddH_2O added to the reaction volume of 10 µl. The cycling conditions were: initial denaturation in 94 °C for 12

minutes; 30 cycles of: denaturation in 94 °C for 30 seconds, primer annealing in 61 °C for 15 seconds and extension in 72 °C for 45 seconds; final extension in 72 °C for 10 minutes and cooling down to 10 °C until further use. The PCR products were pooled together, diluted in 2:75 proportions in Hi-Di formamide and run in a single capillary gel electrophoresis on an ABI 3730xl DNA Analyzer with LIZ-500 size standard (Applied Biosystems). GeneMapper 4.0 software (Applied Biosystems) was used for allele calling (Figure 4).

				PCR product length (bp)	
Deletion	**Forward-primer**	**Label**	**Reverse-primers**	**Normal allele**	**Deleted allele**
DLG2	CAACTGCAATTTC CTTCTGGA	FAM	AGAGTAGAGGCAA GGCAGCA	287	0[a]
			TTTGAAGGGCAGT TTGCAC	33,390[b]	217
ERBB4	GTAAGTCTTGCCC GAAGCTG	VIC	GGAGGTGGGTGTA TTTGTTCC	492	0[a]
			TGTGAGAACAGGC CTTGGA	8613[b]	565
NRG3	GGGGAAATGATT GTGGTTCA	NED	AGGGGTGTGGAGG ATATAGGA	197	0[a]
			AAATGCCTGGATC AAACCAA	12,119[b]	155

[a] Reverse-primer annealing locus deleted.

[b] PCR product too long to be efficiently amplified.

Table 1. PCR primers used in InGoLF-genotyping

Because InGoLF is currently applicable only for deletions, the duplication in *LARGE* was not genotyped in the larger sample set. The "grey zones" in *UTRN* deletion were too large to be efficiently covered with primers needed in the search for the true break points, and thus this deletion was also not genotyped.

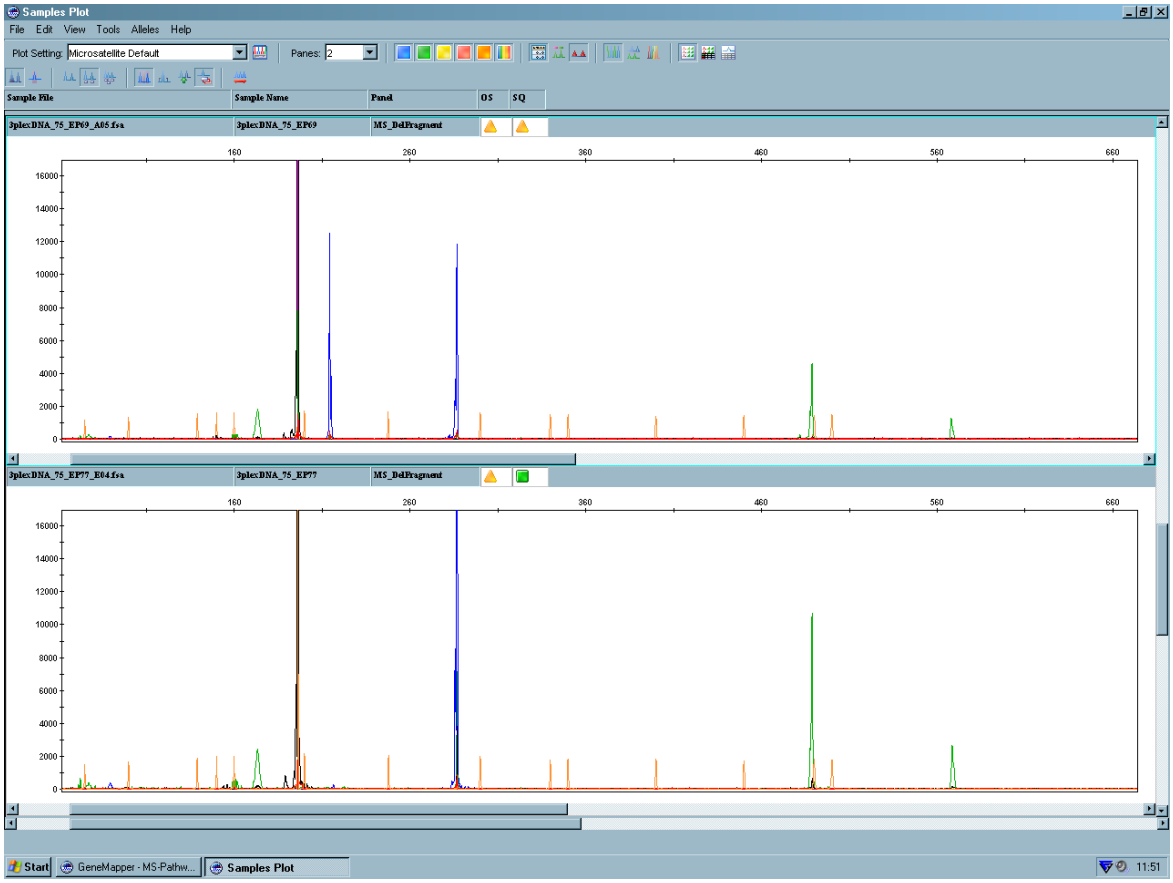

Figure 4. Screen shot from the InGoLF genotyping on GeneMapper 4.0. Window shows the allele peaks for two different samples from ABI 3730xl capillary gel electrophoresis. Both samples are homozygous for the normal allele of *NRG3* (black peak, 197 bp), first sample is heterozygous for the deletion in *DLG2* (blue peaks, 217 bp for the deleted allele and 287 bp for the normal allele), and both samples are heterozygous for the deletion in *ERBB4* (green peaks, 492 bp for the normal allele and 565 bp for the deleted allele). Orange peaks are LIZ-500 size standard marking different PCR product lengths.

Fisher's exact test was used to monitor for statistical difference between the frequencies of the genotyped deletions in the 703 MS cases and 1051 controls. Fisher's test was considered more suitable than the commonly used χ^2-test, as the former allows for data point occurrences of less than 5.

3 Results

We found 106 copy number variable regions in the 68 MS cases with QuantiSNP v1.0 software and visual assurance from the BeadStudio (Appendix 1). Six of the identified CNV regions were novel (not listed in the Database of Genomic Variants (August 5th, 2009 updated version, http://projects.tcag.ca/variation/ (9)) (Table 2). All of these were very rare events and were found only in one individual.

Chr	Start (b35)	End (b35)	Size (bp)	N of SNPs	Copy Number type	Genes
2	54354348	54404322	49975	7	loss	ACYP2 intron, TSPYL6
3	24134391	24172277	37887	8	loss	THRB, last 6 exons
9	21743138	21761241	18104	4	gain	MTAP promoter
13	67542909	67552774	9866	3	gain	
14	73494691	73538366	43676	4	loss	COQ6 last 8 exons, ENTPD5 exons 1-13
18	56949141	56950768	1628	3	loss	

Table 2. Novel CNVs detected in MS cases

We reasoned that the genes mapping to all the 106 CNVs identified in the MS cases could be members of a common pathway involved in the disease etiology. We tested this hypothesis using Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com) to search for connecting pathways for the genes in the CNV regions. Indeed, we found one pathway involving the genes *NRG3*, *ERBB4*, *DLG2*, *UTRN* and *LARGE* (Figure 5). This pathway potentially regulates oligodendrocyte differentiation and myelin sheet formation. Neuregulin 3 (coded by NRG3-gene) is a neural growth factor that inhibits apoptosis of oligodendrocyte precursor cells through its receptor ERBB4 and PI3-K-pathway (17). DLG2 (PSD-93, chapsyn-110) is likely to be involved in the assembly of ERBB4 to the cell membranes, in ERBB4 turnover and neuregulin signaling (18,19). The differentiation of oligodendrocyte precursors and the proper formation of myelin sheets by mature oligodendrocytes seems to be regulated by interactions between neuregulin-ERBB – mediated signaling, laminin-2-integrin signaling and laminin-2-dystroglycan (DG) signaling (20,21). Though the exact intracellular pathways in laminin-2-dystroglycan signaling mediated myelin sheet formation are unknown, dystroglycan is known to bind utrophin (UTRN). Utrophin in turn binds β2-syntrophin (SNTB2), which has a binding site

for ERBB4 (18), thus potentially enabling direct interactions between ERBB4 and dystroglycan-utrophin-complex. Like-glycosyltransferase (LARGE) is an important glycosyl transferase for dystroglycan. Mutations in LARGE-gene cause congenital muscular dystrophy type 1D (MDC1D), a muscular dystrophy with profound mental retardation, white matter changes and subtle structural abnormalities on brain MRI (22).

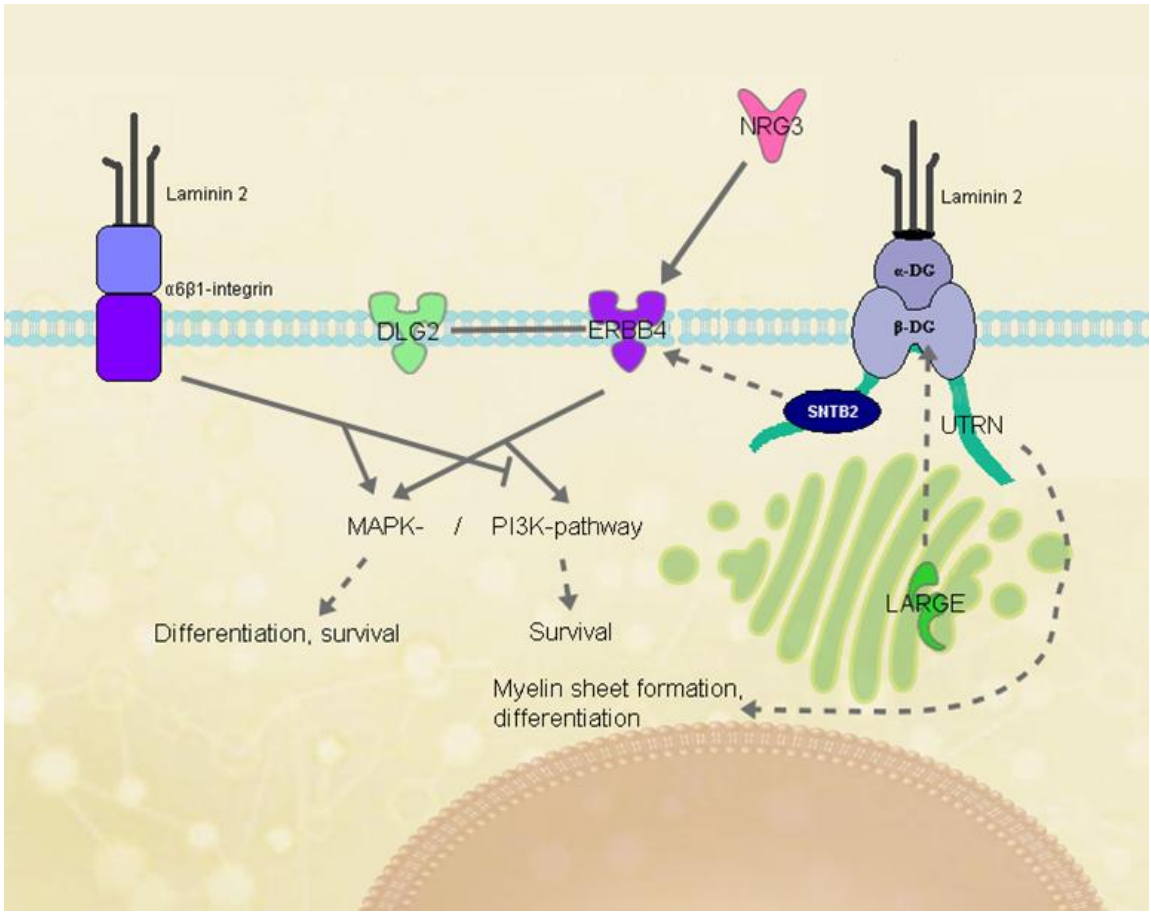

Figure 5. Regulation of oligodendrocyte survival and differentiation. NRG3-ERBB4 enhances oligodendrocyte survival, and together with laminin-2, promotes differentiation. When signaling trough dystroglycan (DG) and possibly utrophin (UTRN), laminin-2 regulates myelin sheet formation. Figure modified from Ingenuity Pathway Analysis (Ingenuity® Systems, www.ingenuity.com). (Figure from Jakkula, *et al*, 2010)

Re-sequencing of the break points for deletions in *ERBB4*, *NRG3* and *DLG2* revealed the true deletions to be in positions chr2:212,892,210-212,900,258, chr10:82,869,397-82,881,361 and chr11:84,215,352-84,248,525, respectively (UCSC Genome Browser, hg18, March 2006 assembly). We used an in-house built method, InGoLF (Instant Genotyping of Lost DNA Fragments) to genotype these deletions in 703 MS cases and 1051 controls from all over Finland. Of the 68 MS cases originally genotyped on Illumina HumanHap300 chip, 48 were also genotyped with InGoLF. All the deletions found with QuantiSNP were also found with InGoLF. We found one *ERBB4* deletion with InGoLF in one of these 48 cases that was not detected with QuantiSNP, giving an overall sensitivity of 100% and specificity of 99% to the method.

Altogether, we found 69 *ERBB4* deletions, 28 *NRG3* promoter deletions and 1 *DLG2* deletion in the 703 MS cases. The found *DLG2* deletion was the same as the one found with QuantiSNP in the original sample set of 68 cases, thus proving this deletion to be a rare, single event. The prevalence of the deletions in the control sample was similar; we found 124 deletions in *ERBB4*, 40 deletions in *NRG3* promoter and none in *DLG2*. Results of the Fisher's exact test for the samples genotyped with InGoLF are given in Table 3. We found no evidence for association of these deletions to MS disease. The *ERBB4* deletion seemed to be as common in both the cases and controls originating from the high risk MS region in Southern Ostrobothnia as in elsewhere of Finland (frequency of 0.10 in cases and 0.11 in controls in Southern Ostrobothnia versus 0.11 in cases and 0.11 in controls from rest of Finland), whereas the *NRG3* promoter deletion was slightly more common in Southern Ostrobothnia (0.05 and 0.06 versus 0.04 and 0.03, respectively).

Deletion	**Cases (n=703)**		**Controls (n=1051)**		**Fisher's Exact test**
	losses	loss frequency	losses	loss frequency	*p*-value
ERBB4	69	0,11	124	0,12	0,388
NRG3	28	0,04	40	0,04	0,900
DLG2	1	0,00	0	0,00	0,404

Table 3. Results of the InGoLF-genotyping of the three pathway deletions

4 Discussion

The two identified extended pedigrees with multiple affected MS patients and common ancestry support the hypothesis of a founder effect in the Sothern Ostrobothnia. Such founder effect could lead to enrichment of rare, deleterious mutations leading to the increased disease risk in the isolate. However, the novel CNVs identified in these distantly related MS patients were all seen only in one individual, and are likely not causative for the disease. We did not find a single common CNV for all or even the majority of the studied MS patients. Given the complex nature of the disease inheritance, this was anticipated.

As no single CNV could be identified to be associated with MS in the distantly related samples, a combination of CNVs affecting genes on same pathway could result in increased disease risk. Our pathway analyses revealed a highly interesting conduit, with multiple genes affected by a CNV, involved in oligodendrocyte development and survival. Though most of the studies considering the pathway described are performed with cells derived from developing murine brain, similar interactions could happen in demyelinated white matter during remyelination, and altered protein function could plausibly predispose to MS. However, no significant increase in the frequency of these deletions were identified when we genotyped the *ERBB4, DLG2* and *NRG3* deletions in a larger sample set of cases and controls. Although no association was detected, there was a slight difference in the frequency of *NRG3* deletion between the Southern Ostrobothnian samples and samples from the rest of Finland. Albeit this may not affect to the increased MS disease risk in the subisolate, it emphasizes that similar caution must be taken when studying CNVs as with SNP studies in populations with internal substructures (23).

The Illumina HumanHap300 chip is not ideal for copy number variation estimation. Its marker density is relatively sparse, and thus only the large CNVs can be detected. The mean length of the CNVs we found was over 190 Kb, whereas the mean length of CNVs detected on more dense arrays has been less than 27 Kb (24). Moreover, the genomic

regions with SNPs that have not passed the conventional quality control steps, such as regions with inversions or common duplications and deletions, are underrepresented on the Illumina HumanHap300 chip. Thus, SNPs on the array tend to be selected from regions without common CNVs. Taking all this to account, we were likely to miss a huge portion of common copy number variations. By contrast, we were able to find six new CNVs and many rare CNVs. To truly evaluate the effect of rare genetic variations on disease risk huge sample sizes and large collaborations are necessary, and thus we could not profoundly assess the role of the identified rare CNVs to MS disease in this study. However, such large sample collection is impossible from a small subisolate. In an isolated population with a founder effect, rare family specific mutations leading to the increased disease risk should be considered as well. These cannot be identified with genotyping assays designed for common variations. Sequencing of whole genomes of the families from Southern Ostrobothnia could enlighten the causes of the increased MS disease risk within this region.

References

(1) Soinila S, Kaste M, Somer H, Alaranta H. Neurologia. 2. revised ed. Helsinki: Duodecim; 2006.

(2) Sumelahti ML, Tienari PJ, Wikstrom J, Palo J, Hakama M. Regional and temporal variation in the incidence of multiple sclerosis in Finland 1979-1993. Neuroepidemiology 2000 Mar-Apr;19(2):67-75.

(3) International Multiple Sclerosis Genetics Consortium, Hafler DA, Compston A, Sawcer S, Lander ES, Daly MJ, et al. Risk alleles for multiple sclerosis identified by a genomewide study. N Engl J Med 2007 Aug 30;357(9):851-862.

(4) International Multiple Sclerosis Genetics Consortium (IMSGC). Refining genetic associations in multiple sclerosis. Lancet Neurol 2008 Jul;7(7):567-569.

(5) De Jager PL, Jia X, Wang J, de Bakker PI, Ottoboni L, Aggarwal NT, et al. Meta-analysis of genome scans and replication identify CD6, IRF8 and TNFRSF1A as new multiple sclerosis susceptibility loci. Nat Genet 2009 Jul;41(7):776-782.

(6) International Multiple Sclerosis Genetics Consortium (IMSGC). The expanding genetic overlap between multiple sclerosis and type I diabetes. Genes Immun 2009 Jan;10(1):11-14.

(7) Jakkula E, Leppa V, Sulonen AM, Varilo T, Kallio S, Kemppinen A, et al. Genome-wide association study in a high-risk isolate for multiple sclerosis reveals associated variants in STAT3 gene. Am J Hum Genet 2010 Feb 12;86(2):285-291.

(8) International Multiple Sclerosis Genetics Consortium, Wellcome Trust Case Control Consortium 2, Sawcer S, Hellenthal G, Pirinen M, Spencer CC, et al.

Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 2011 Aug 10;476(7359):214-219.

(9) Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet 2004 Sep;36(9):949-951.

(10) Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, et al. Global variation in copy number in the human genome. Nature 2006 Nov 23;444(7118):444-454.

(11) McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, et al. Common deletion polymorphisms in the human genome. Nat Genet 2006 Jan;38(1):86-92.

(12) Poser CM, Paty DW, Scheinberg L, McDonald WI, Davis FA, Ebers GC, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. Ann Neurol 1983 Mar;13(3):227-231.

(13) Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res 2007;35(6):2013-2025.

(14) Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 2007 November 1;17(11):1665-1674.

(15) Sulonen AM, Kallio SP, Ellonen P, Suvela M, Elovaara I, Koivisto K, et al. No evidence for shared etiology in two demyelinative disorders, MS and PLOSL. J Neuroimmunol 2009 Jan 3;206(1-2):86-90.

(16) Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 2007 Jul;35(Web Server issue):W71-4.

(17) Carteron C, Ferrer-Montiel A, Cabedo H. Characterization of a neural-specific splicing form of the human neuregulin 3 gene involved in oligodendrocyte survival. J Cell Sci 2006 Mar 1;119(Pt 5):898-909.

(18) Garcia RA, Vasudevan K, Buonanno A. The neuregulin receptor ErbB-4 interacts with PDZ-containing proteins at neuronal synapses. Proc Natl Acad Sci U S A 2000 Mar 28;97(7):3596-3601.

(19) Huang YZ, Won S, Ali DW, Wang Q, Tanowitz M, Du QS, et al. Regulation of neuregulin signaling by PSD-95 interacting with ErbB4 at CNS synapses. Neuron 2000 May;26(2):443-455.

(20) Colognato H, Baron W, Avellana-Adalid V, Relvas JB, Baron-Van Evercooren A, Georges-Labouesse E, et al. CNS integrins switch growth factor signalling to promote target-dependent survival. Nat Cell Biol 2002 Nov;4(11):833-841.

(21) Colognato H, Galvin J, Wang Z, Relucio J, Nguyen T, Harrison D, et al. Identification of dystroglycan as a second laminin receptor in oligodendrocytes, with a role in myelination. Development 2007 May;134(9):1723-1736.

(22) Longman C, Brockington M, Torelli S, Jimenez-Mallebrera C, Kennedy C, Khalil N, et al. Mutations in the human LARGE gene cause MDC1D, a novel form of congenital muscular dystrophy with severe mental retardation and abnormal glycosylation of alpha-dystroglycan. Hum Mol Genet 2003 Nov 1;12(21):2853-2861.

(23) Jakkula E, Rehnstrom K, Varilo T, Pietilainen OP, Paunio T, Pedersen NL, et al. The genome-wide patterns of variation expose significant substructure in a founder population. Am J Hum Genet 2008 Dec;83(6):787-794.

(24) McCarroll SA, Kuruvilla FG, Korn JM, Cawley S, Nemesh J, Wysoker A, et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. Nat Genet 2008 Oct;40(10):1166-1174.

Appendix 1

Description of all CNVs found in the MS samples. Table originally published in Jakkula, *et al*, 2010.

Chr	Start (b35)	End (b35)	Size (bp)	N of SNPs	Copy Number type	Cases (n=68)	Cases freq	Db of Genomic Variants, August 5th 2009 Freeze	Genes
1	2058523	2312823	254301	21	gain	3	4.4%	Known variation	PRKCZ, C1orf86, SKI, MORN1
1	12354566	12763485	408920	36	gain	1	1.5%	Known variation	VPS13D, DHRS3, AADACL4, AADACL3, C1orf158, after TNFRSF1B
1	102379409	102651711	272303	21	loss	1	1.5%	Known variation	
1	120776649	120923841	147193	7	gain	2	2.9%	Known variation	
1	193569717	193628745	59029	6	loss	4	5.9%	Known variation	CFHR4
2	19443	189678	170236	14	gain	1	1.5%	Known variation	FAM110C
2	35725730	35999293	273564	40	loss	1	1.5%	Known variation	
2	49511191	49664240	153050	21	loss	2	2.9%	Known variation	
2	53071353	53103920	32568	4	loss	1	1.5%	Known variation	
2	54354348	54404322	49975	7	loss	1	1.5%	NEW	ACYP2 intron, TSPYL6
2	86983132	87615568	632437	3	gain	1	1.5%	Known variation	CD8B1, PLGLB1, RGPD1
2	89772948	89932893	159946	3	loss	7	10.3%	Known variation	
2	99215459	99306408	90950	4	gain	3	4.4%	Known variation	TSGA10, C2orf15, LIPT1, MITD1, MRPL30, LYG2, LYG1
2	110214618	110292098	77481	7	loss	1	1.5%	Known variation	MALL, NPHP1
2	212981040	213018166	37127	5	loss	3	4.4%	Known variation	ERBB4 intron
2	241106342	241131279	24938	6	gain	1	1.5%	Known variation	GPC1
3	41894	108412	66519	8	gain	1	1.5%	Known variation	CHL1 promoter
3	4021691	4329697	308007	35	loss	1	1.5%	Known variation	SETMAR
3	5383302	5411345	28044	6	loss	1	1.5%	Known variation	
3	24134391	24172277	37887	8	loss	1	1.5%	NEW	THRB, last 6 exons
3	65166887	65187636	20750	3	loss	1	1.5%	Known variation	
3	152993783	153028739	34957	7	loss	3	4.4%	Known variation	AADAC
4	84915835	84969552	53718	10	loss	1	1.5%	Known variation	

4	132303980	132685819	381840	24	gain	1	1.5%	Known variation	
4	167425804	167482371	56568	8	gain	1	1.5%	Known variation	After TLL1
4	190200031	191131631	931601	93	gain	1	1.5%	Known variation	
5	19055301	19305295	249995	22	loss	1	1.5%	Known variation	
5	97074222	97121798	47577	7	loss	5	7.4%	Known variation	RIOK2 near
5	104465860	104510644	44785	3	loss	1	1.5%	Known variation	
5	178661436	178854467	193032	25	gain	1	1.5%	Known variation	ADAMTS2
6	5496231	5522147	25917	7	loss	1	1.5%	Known variation	FARS2, intron
6	29076909	29287216	210308	25	loss	1	1.5%	Known variation	ZNF311, OR2W1, OR2B3, OR2J3, OR2J2
6	67044129	67104015	59887	11	loss	5	7.4%	Known variation	BAI3 promoter
6	95472452	95649511	177060	9	loss	1	1.5%	Known variation	
6	137942200	138003365	61166	21	gain	1	1.5%	Known variation	OLIG3 promoter, TNFAIP3 promoter
6	144743809	145071977	328169	32	loss	1	1.5%	Known variation	UTRN, many exons
7	3175715	3252673	76959	8	loss	1	1.5%	Known variation	SDK1, intron
7	6643875	7123923	480049	54	gain	1	1.5%	Known variation	C1GALT1
7	8882102	8996708	114607	23	loss	3	4.4%	Known variation	After NXPH1
7	9790639	9877397	86759	9	loss	2	2.9%	Known variation	
7	12560343	12685347	125005	17	loss	1	1.5%	Known variation	
7	69761016	70029858	268843	22	gain	1	1.5%	Known variation	After AUTS2, WBSCR17 promoter
7	70919613	71427575	507963	52	gain	1	1.5%	Known variation	CALN1, exons 1-3
7	75690975	76059191	368217	6	gain	1	1.5%	Known variation	ZP3, DTX2, UPK3B, POMZP3
7	119899100	119906697	7598	3	loss	1	1.5%	Known variation	KCND2, intron
7	152959823	153145033	185211	20	gain	1	1.5%	Known variation	DPP6
8	5590045	5591685	1641	3	loss	1	1.5%	Known variation	
8	16170358	16306880	136523	11	loss	1	1.5%	Known variation	MSR1 promoter
8	87256166	87403084	146919	11	gain	1	1.5%	Known variation	SLC7A13, WWP1 promoter
8	92192384	92243291	50908	4	loss	5	7.4%	Known variation	Hypothetical protein FLJ27355, after OTUD6B, SLC26A7 promoter
8	137757412	137919630	162219	13	loss	1	1.5%	Known variation	
9	5296824	5325470	28647	6	loss	1	1.5%	Known variation	RLN1
9	10651370	10676202	24833	7	loss	1	1.5%	Known variation	PTPRD promoter
9	21743138	21761241	18104	4	gain	1	1.5%	NEW	MTAP promoter

9	28534375	28556849	22475	5	loss	3	4.4%	Known variation	LINGO2
9	69329605	69348516	18912	4	loss	1	1.5%	Known variation	APBA1, intron
9	70946694	71169744	223051	32	gain	1	1.5%	Known variation	TRPM3 promoter & exon1
9	71310230	71589788	279559	48	gain	1	1.5%	Known variation	TMEM2
10	15030375	15100889	70515	6	loss	1	1.5%	Known variation	DCLRE1C, MEIG1
10	47013328	47173619	160292	12	gain	12	17.6%	Known variation	ANXA8 promoter
10	51457810	51805221	347412	28	gain	1	1.5%	Known variation	FAM21A, ASAH2, SGMS1
10	66102105	67710331	1608227	209	loss	1	1.5%	Known variation	CTNNA3
10	81567594	82006206	438613	40	gain	1	1.5%	Known variation	SFTPD, C10orf57, PLAC9, ANXA11
10	82869699	82875955	6257	3	loss	2	2.9%	Known variation	NRG3 promoter
11	38188336	38965147	776812	58	loss	1	1.5%	Known variation	
11	84219051	84245672	26622	4	loss	1	1.5%	Known variation	DLG2, intron
11	89843914	91455249	1611336	121	loss	1	1.5%	Known variation	
11	133873030	134225383	352354	69	gain	1	1.5%	Known variation	
12	7884583	8017012	132430	14	gain	5	7.4%	Known variation	SLC2A14, SLC2A3
12	31157554	31293957	136404	12	gain	1	1.5%	Known variation	OVOS2
12	31898373	31954269	55897	12	gain	2	2.9%	Known variation	
12	42212399	42288535	76137	12	loss	2	2.9%	Known variation	ADAMTS20
12	81671084	81707620	36537	5	loss	1	1.5%	Known variation	TMTC2, intron
12	98788860	98966181	177322	18	gain	1	1.5%	Known variation	ANKS1B exon1, KIAA0701 last 7 exons
12	130255197	130339814	84618	11	loss	1	1.5%	Known variation	
13	67542909	67552774	9866	3	gain	1	1.5%	NEW	
13	83009774	83055928	46155	8	loss	8	11.8%	Known variation	After SLITRK1
13	94733590	94779857	46268	11	gain	1	1.5%	Known variation	ABCC4 promoter & exon1
13	94797740	94825508	27769	7	gain	1	1.5%	Known variation	
14	73494691	73538366	43676	4	loss	1	1.5%	NEW	COQ6 last 8 exons, ENTPD5 exons 1-13
15	20347960	20777695	429736	50	gain	1	1.5%	Known variation	TUBGCP5, CYFIP1, NIPA2, NIPA1
15	38624662	39476524	851863	53	loss	2	2.9%	Known variation	CCDC32, RPUSD2, CASC5, RAD51, FAM82C, GCHFR, DNAJC17, ZFYVE19, PPP1R14D, SPINT1, RHOV, VPS18, DLL4, CHAC1, INOC1, EXDL1, CHP, OIP5, NUSAP1, NDUFAF1
15	98736980	98770528	33549	5	loss	1	1.5%	Known variation	LASS3

16	21515973	21647775	131803	12	loss	1	1.5%	Known variation	METTL9, IGSF6, OTOA
16	81289305	81385638	96334	27	loss	1	1.5%	Known variation	CDH13, intron
17	30708148	30792312	84165	14	loss	1	1.5%	Known variation	SLFN11 promoter & exons 1-4, SLFN12, SLFN13 exon1
17	38782929	38958044	175116	13	gain	1	1.5%	Known variation	ARL4D, DHX8, near BRCA1
17	74878104	74905197	27094	9	gain	1	1.5%	Known variation	HRNBP3
18	1917798	1970668	52871	9	loss	1	1.5%	Known variation	
18	56949141	56950768	1628	3	loss	1	1.5%	NEW	
18	62947038	63159299	212262	31	gain	1	1.5%	Known variation	After DSEL
18	64897188	64909977	12790	3	loss	5	7.4%	Known variation	After hypothetical proteins CCDC102B & TXNDC10. DOK6 promoter
19	3755701	3786177	30477	9	gain	1	1.5%	Known variation	KIAA1086, ATCAY promoter, MATK promoter
19	7658063	7691713	33651	6	gain	1	1.5%	Known variation	FCER2 whole gene & promoter
19	48066441	48350666	284226	6	loss	2	2.9%	Known variation	PSG1, PSG6, PSG7, PSG11, PSG2, PSG5, PSG4, PSG9
19	59994795	60069820	75026	5	loss	1	1.5%	Known variation	KIR3DP1, KIR2DL4, KIR3DL1, KIR2DS4, KIR3DL2, FCAR & NCR1 promoter
20	5219710	5345328	125619	22	loss	1	1.5%	Known variation	PROKR2 whole gene and promoter
20	14874333	15089959	215627	38	loss	1	1.5%	Known variation	MACROD2
21	18981497	19000295	18799	6	gain	1	1.5%	Known variation	PRSS7 promoter
21	43646295	43663581	17287	7	gain	1	1.5%	Known variation	SNF1LK
22	20654301	20920813	266513	48	gain	1	1.5%	Known variation	TOP3B, IGLC1
22	23974960	24235221	260262	10	gain	2	2.9%	Known variation	LRP5L
22	32490635	32503091	12457	6	gain	1	1.5%	Known variation	LARGE, intron
X	6317902	7961924	1644023	92	gain	1	1.5%	Known variation	HDHD1A, STS, VCX, PNPLA4, VCX2
X	91173351	91262603	89253	5	loss	1	1.5%	Known variation	PCDH11X intron & exon3
X	140076396	140280802	204407	30	loss	2	2.9%	Known variation	SPANXA2 intron